

ABSTRACT OF THE DISCLOSURE

5 A system, method and computer program product for  
determining whether a test sample is in a first or a second  
class of data (for example: cancerous or normal), comprising:  
extracting a plurality of emerging patterns from a training  
data set, creating a first and second list containing  
10 respectively, a frequency of occurrence of each emerging  
pattern that has a non-zero occurrence in the first and in the  
second class of data; using a fixed number of emerging  
patterns, calculating a first and second score derived  
respectively from the frequencies of emerging patterns in the  
first list that also occur in the test data, and from the  
15 frequencies of emerging patterns in the second list that also  
occur in the test data; and deducing whether the test sample  
is categorized in the first or the second class of data by  
selecting the higher of the first and the second score.

20

25

30

35